Stereoscopic First Person View System for Drone Navigation

Nikolai Smolyanskiy*†‡ and Mar Gonzalez-Franco*†

Microsoft Research, One Microsoft Way, Redmond, WA, USA

Ground control of unmanned aerial vehicles (UAV) is a key to the advancement of this technology for commercial purposes. The need for reliable ground control arises in scenarios where human intervention is necessary, e.g. handover situations when autonomous systems fail. Manual flights are also needed for collecting diverse datasets to train deep neural network-based control systems. This axiom is even more prominent for the case of unmanned flying robots where there is no simple solution to capture optimal navigation footage. In such scenarios, improving the ground control and developing better autonomous systems are two sides of the same coin. To improve the ground control experience, and thus the quality of the footage, we propose to upgrade onboard teleoperation systems to a fully immersive setup that provides operators with a stereoscopic first person view (FPV) through a virtual reality (VR) head-mounted display. We tested users ($n = 7$) by asking them to fly our drone on the field. Test flights showed that operators flying our system can take off, fly, and land successfully while wearing VR headsets. In addition, we ran two experiments with prerecorded videos of the flights and walks to a wider set of participants ($n = 69$ and $n = 20$) to compare the proposed technology to the experience provided by current drone FPV solutions that only include monoscopic vision. Our immersive stereoscopic setup enables higher accuracy depth perception, which has clear implications for achieving better teleoperation and unmanned navigation. Our studies show comprehensive data on the impact of motion and simulator sickness in case of stereoscopic setup. We present the device specifications as well as the measures that improve teleoperation experience and reduce induced simulator sickness. Our approach provides higher perception fidelity during flights, which leads to a more precise better teleoperation and ultimately translates into better flight data for training deep UAV control policies.

Keywords: drone, virtual reality, stereoscopic cameras, real time, robots, first person view

INTRODUCTION

One of the biggest challenges of unmanned aerial vehicles (UAV) (drones, UAV) is their teleoperation from the ground. Both researchers and regulators are pointing at mixed scenarios in which control of autonomous flying robots creates handover situations where humans must regain control under specific circumstances (Russell et al., 2016). UAV fly at high speeds and can be dangerous. Therefore, ground control handover needs to be fast and optimal. The quality of the teleoperation has implications not only for stable drone navigation but also for developing autonomous systems themselves because captured flight footage can be used to train new control policies. Hence, there

is a need to better understand the implications of the ground control and find a solution that provides the best combination of human teleoperation and autonomous control in one visual setup. Existing commercial drones can be controlled either from a third person view (TPV) or from a first person view (FPV). The TPV type of operation creates control and safety problems, since operators have difficulty guessing the drone orientation and may not be able to keep eye contact with them even at short distances (>100 m). Furthermore, given that effective human stereo vision is limited to 19 m (Allison et al., 2009), it is hard for humans to estimate drone's position relative to nearby objects to avoid obstacles and control attitude/altitude. In such cases, the ground control gets more challenging the further the drone gets from the operator.

Typically, more advanced consumer drone control setups include FPV goggles with VGA (640 × 480) or WVGA (800 × 480) resolution and limited field of view (FOV) (30–35°).[1] Simpler systems may use regular screens or smartphones for displaying FPV. Most existing systems only send monocular video feed from the drone's camera to the ground. Examples of these systems include FatShark, Skyzone, and Zeiss.[2] These systems are able to transmit video feeds over analog signals in 900 MHz, 1.2–1.3 GHz, and 2.4 GHz bands with low latency video transmission and graceful reception degradation, allowing flying drones at high speeds and near obstacles. These systems can be combined with FPV glasses: e.g., Parrot Bebop or DJI drones with Sony or Zeiss Cinemizer FPV glasses (see text footnote 2). Current drone racing championship winners are flying on these precise setups and achieve average speeds of 30 m/s. Unfortunately, analog and digital systems proposed to date have a very narrow (30–40°) FOV and low video resolution and frame rates, cannot be easily streamed over Internet, and are limited to a monocular view (which affects distance estimations). These technical limitations have been shown to significantly affect the performance and experience in robotic telepresence in the past (Chen et al., 2007). In particular, sensory fidelity is affected by head-based rendering (produced by head tracking and digital/mechanical panning latencies), frame rate, display size, and resolution (Bowman and McMahan, 2007). Experiments involving robotic telecontrol and task performance have shown that FPV stereoscopic viewing presents significant advantages over monocular images; tasks are performed faster and with fewer errors under stereoscopic conditions (Drascic, 1991; Chen et al., 2007). In addition, providing FPV with head tracking and head-based rendering also affects task performance regardless of stereo viewing; e.g., by using head tracking even with simple monocular viewing participants improved at solving puzzles under degraded image resolutions (Smets and Overbeeke, 1995). Even though depth estimation can be performed using monocular motion parallax through controlled camera movements (Pepper et al., 1983), the variation of the intercamera distance affects performance on simple depth matching tasks involving peg alignment (Rosenberg, 1993). Indeed, a separation of only 50% (~3.25 cm) between stereoscopic cameras significantly

improved performance over conventional 2D monoscopic laparoscopes (Fishman et al., 2008), all in all, showing clear evidence of significant performance implications between monoscopic and stereoscopic visions for telerobot operation and potentially for FPV drone control.

In sum, most of the technical factors described as basic for an effective teleoperation of robots in FPV are related with having a more immersive experience (wider FOV, stereo vs. mono, head tracking, image quality, etc.), which can be ultimately achieved through virtual reality (VR) (Spanlang et al., 2014). In that line, there have been several attempts to use new VR headsets like Oculus Rift for drone FPV control (Hayakawa et al., 2015; Ramasubramanian, 2015). Oculus Rift VR headset may provide 360° view with help of digital panning, potentially creating an immersive flight experience given enough camera data. However, it is known that VR headsets and navigation, which does not stimulate the vestibular system, generate simulator sickness (Pausch et al., 1992). This is increased by newer head-mounted displays (HMDs) with greater FOV, as simulator sickness increases with wider FOVs and with navigation (Pausch et al., 1992). Hence, simulator sickness has become a serious challenge for the advancement of VR for teleoperation of drones.

Some experiments have tried to reduce simulator sickness during navigation by reducing the FOV temporarily (Fernandes and Feiner, 2016) or increasing the sparse flow on the peripheral view also affecting the effective FOV (Xiao and Benko, 2016). Postural stability can also modulate the prevalence of sickness, being more under control when participants are sitting rather than standing (Stoffregen et al., 2000). In general, VR navigation induces mismatched motion, which is internally related to incongruent multisensory processing in the brain (González Franco, 2014; Padrao et al., 2016). Indeed, to provide a coherent FPV in VR, congruent somatosensory stimulation and small latencies are needed, i.e., as the head rotates so does the rendering (Sanchez-Vives and Slater, 2005; Gonzalez-Franco et al., 2010). If the tracking systems are not responsive enough or induce strong latencies, the VR experience will be compromised (Spanlang et al., 2014), and this alone can induce simulator sickness.

Some drone setups have linked the head tracking to mechanically panned cameras control[3] to produce coherent somatosensory simulation. Such systems provide real-time, stereo FPV from the drone's perspective to the operator (Hals et al., 2014), as shown by the SkyDrone FPV.[4] However, the latency of camera panning inherent in the mechanical systems can induce significant simulator sickness (Hals et al., 2014). Head tracking has also been proposed to directly control the drone navigation and rotations with attached fixed cameras; however, those systems also induced severe lag and motion sickness (Mirk and Hlavacs, 2015), even if a similar approach might work in other robotic scenarios without navigation (Kishore et al., 2014). Overall, latencies in panning can be overcome with digital panning of wide video content captured by wide-angle cameras. This approach has been already implemented using only monoscopic view in commercial drones

---

[1] FatShark: http://www.fatshark.com/.
[2] FPV guide: http://www.dronethusiast.com/the-ultimate-fpv-system-guide/.

[3] Oculus FPV: https://github.com/Matsemann/oculus-fpv.
[4] Sky Drone: http://www.skydrone.aero/.

such as the Parrot Bebop drone, without major simulator sickness being induced. However, still there is no generic option for stereoscopic FPV control for drone teleoperation nor it is known the extent of the associated simulator sickness. A solution with stereoscopic video feed would have clear implications not only for operators' depth perception but also can help with autonomous drone navigation and path planning (Zhang et al., 2013). UAV equipped with mono-camera-based need to rely on frame accumulation to estimate depth and ego motion with mono SLAM/visual odometry (Lynen et al., 2013; Forster et al., 2014). Unfortunately, these systems cannot recover world scale from visual cues alone and rely on IMUs with gyroscopes, accelerometers (Lynen et al., 2013), or special calibration procedures to estimate world scale. Inaccuracies in scale estimation contribute to errors in vehicle pose computation. In addition, depth map estimation requires accurate pose information for each frame, which is challenging for mono SLAM systems. These issues have led several drone makers to consider using stereo systems[5,6] that can provide world scale and accurate depth maps immediately from their stereo cameras only for autonomous control. However, current stereo systems typically have wide camera baselines[7] to increase depth perception and are not suitable immersive stereoscopic FPV for human operators. Hence, they are not suitable for optimal handover situations. It would be beneficial to have one visual system that would allow for both immersive teleoperation and can be used in autonomous navigation.

For both cases of ground controlled and autonomous flying robots, there is a need to better understand the implications of the FPV control and to find a solution that will provide both the best operator experience and the best conditions for autonomous control. Therefore, we propose an immersive stereoscopic teleoperation system for drones that allows high-quality ground control in FVP, improves autonomous navigation, and provides better capabilities for collecting video footage for training future autonomous and semiautonomous control policies. Our system uses a custom-made wide-angle stereo camera mounted on a commercial hexacopter drone (DJI S900). The camera streams all captured video data digitally (as H.264 video feed) to the ground station connected to a VR headset. Given the preliminary importance of stereoscopic view on other teleoperation scenarios, we hypothesize important implications of stereoscopic view also for the specific case of ground drone control in FPV. We tested the prototype in the wild by asking users ($n = 7$) to fly our drone on the field. Also we ran two experiments with prerecorded videos of the flights and walks with a wider set of participants ($n = 69$ and $n = 20$) to compare the current commercial drone FPV solution monoscopic vs. our immersive stereoscopic setup. We measured the quality of the experience, the distance estimation abilities, and simulator sickness. In this article, we present the device and the evidence of how our system improves the ground control experience.

---

[5]DJI Phantom 4: https://www.dji.com/phantom-4.
[6]Parrot´s stereo devkit: https://www.engadget.com/2016/09/08/parrot-drone-dev-kit/.
[7]ZED camera: https://www.stereolabs.com/.

MATERIALS AND METHODS

Apparatus

Our system consists of a custom wide-angle stereo camera with fisheye (185°) lenses driven from an onboard computer that streams the video feed and can be attached to a drone. In our test scenario, we attached our vision system to a modified commercial hexacopter (DJI S900) typically used for professional film making[8] (**Figure 1**).

We chose this drone since it can lift rather heavy camera setup (1.5 kg) and has superior air stability for better video capture. It is possible to miniaturize our setup with better/smaller hardware and mount it on a smaller drone. The 180° left and right stereo frames are fused into one image and streamed live to Oculus Rift DK2 headset as H264 stream. Correct stereo views are rendered on the ground station at 75 FPS based on operator's head poses. This allows the operator to freely look around with little latency in the panning (<10 ms) while flying the drone *via* an RC controller (Futaba) or Xbox controller. **Figure 2** describes our setup.

The stereo camera system is made of two Grasshopper 3 cameras with fisheye Fujinon lenses (2.7 mm focal length and 185° FOV) mounted parallel to each other and separated by 65 mm distance [average human interpupillary distance (IPD)] (Pryor, 1969). Each camera has 2,048 × 2,048 resolution and 1″ CMOS sensor with a global shutter and are synchronized within 10 ms range. In our test setup, the camera is rigidly attached to the DJI S900 hexacopter (without mechanical gimbal), so the pilot can use horizon changes as orientation cues (**Figure 3**). We used simple rubber suspension to minimize vibration noise.

An onboard computer, NVIDIA Tegra TK1, reads left and right video frames from the cameras, concatenates them as one image (stacks on vertically), encodes the fused image in H264 with a hardware accelerated encoder, and then streams it with UDP over WiFi. We used gstreamer1.0 pipeline for streaming and encoding. In normal functioning, the system can reach 28–30 frames per second encoding speed at 1,600 × 1,080 resolution per camera/eye. The onboard computer (Tegra TK1) is also capable of computing depth maps from incoming stereo frames for obstacle avoidance. Unfortunately, it could not do both tasks (video streaming and depth computation) in real time, so our experiments with depth computation were limited and not real time. Better compute (e.g., with newest Tegra TX1) would allow both FPV teleoperation and autonomous obstacle avoidance systems to run simultaneously.

On the ground, the client station provides FPV control of the drone through an Oculus Rift DK2 VR headset attached to a laptop (**Figure 3**) and Xbox or RC controller. In our setup, the digital panning is enabled by the head tracking of the VR headset. We render correct views for left and right eyes based on current user's head pose and by using precomputed UV coordinates to sample the pixels. We initialize these UV coordinates by projecting vertices of two virtual spheres onto fisheye frames using camera intrinsic parameters calculated as in Eq. 1 and then interpolating UVs between neighboring vertices. This initialization process

---

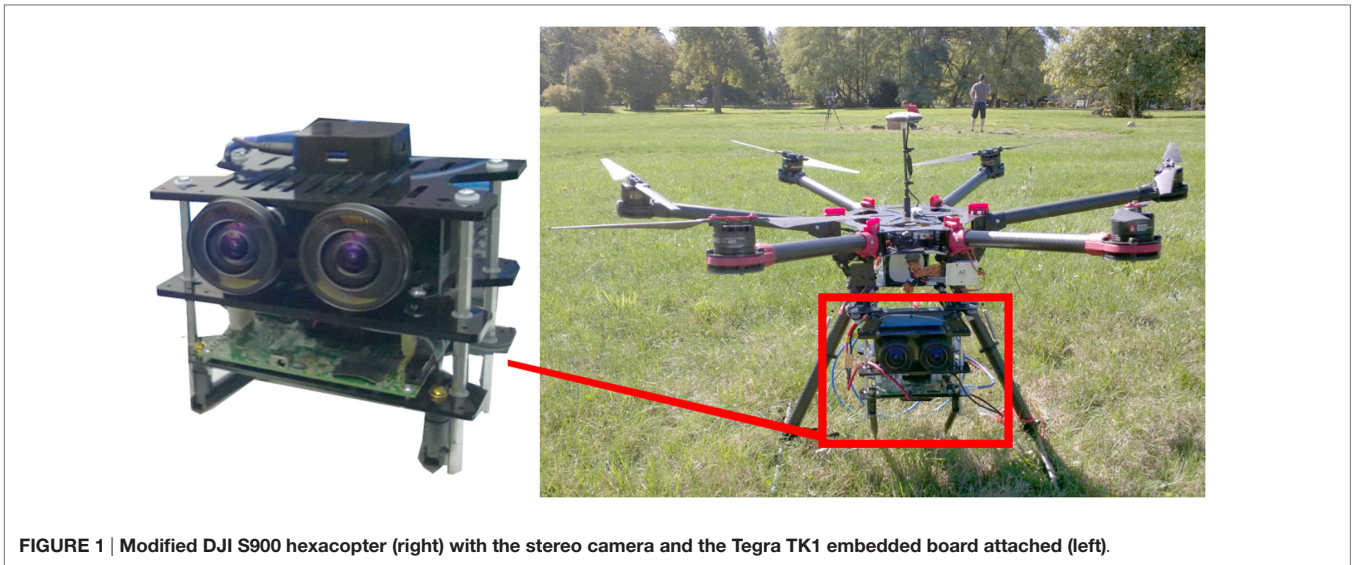[8]DJI S900 hexacopter: http://www.dji.com/product/spreading-wings-s900.

FIGURE 1 | Modified DJI S900 hexacopter (right) with the stereo camera and the Tegra TK1 embedded board attached (left).
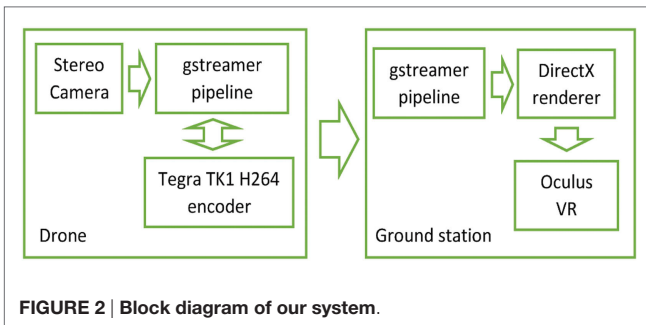

FIGURE 2 | Block diagram of our system.

creates sampling maps that we use during frame rendering. The calibration of the fisheye equidistant camera model intrinsic and extrinsic parameters was done with 0.25 pixels reprojection error by using OpenCV (Zhang, 2000).

Pinhole projection:

$$a = \frac{x}{z}, b = \frac{y}{z}$$
$$r^2 = a^2 + b^2$$
$$\theta = atan(r).$$ (1)

Fisheye distortion:

$$\theta_d = \theta\left(1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + k_4\theta^8\right)$$
$$x' = \left(\frac{\theta_d}{r}\right)a, y' = \left(\frac{\theta_d}{r}\right)b.$$

At runtime, DirectX pipeline only needs to sample pixels from left and right frames per precomputed UV coordinate mappings to render correct views based on viewer's head pose. This process is very fast and allows rendering at 75 frames per second, which avoids judder and reduces simulator sickness even though the incoming video feed is only 30 frames per second. The wide FOV of the video (180°) enables the use of digital panning on the client side, so the operator can look around while wearing Oculus VR headset. Digital panning has less than 10 ms latency and therefore is superior to mechanical panning in terms of reducing simulator sickness. In our setup, the digital panning is enabled by the head tracking of the VR headset and rendered by sampling according to precomputed maps. The mapping is computed *via* our fisheye camera model at initialization, and it projects the hemispheres 3D vertices into the correct UV coordinates on the left–right frames. UV coordinates for 3D points that lie in between those vertices are interpolated by the rendering pipeline (*via* bilinear interpolation). When the head pose changes, the 3D pipeline updates its viewing frustum and only renders pixels that are visible (according to a new head pose).

In total, the system has 250 ± 30 ms latency from the action execution till the video came through the HMD. Most of the delay was introduced during the H264 encoding, streaming, and decoding of the video feed and can be improved with more powerful onboard computing. Nevertheless, the system was usable despite these latency limitations.

Experimental Procedure

Participants ($n = 7$) were asked to fly the drone; in addition, a video stream from the live telepresence setup was recorded in the same open field. Real flight testing was limited due to US Federal Aviation Administration restrictions.

To further study the role of different components on the experience such as the monoscopic vs. stereoscopic view, the perceived quality, and the simulator sickness levels, we also run two additional experiments ($n = 69$ and $n = 20$) using prerecorded videos. In both cases, we provided offline experience of prerecorded videos of a flight and a ground walk. Participants in these conditions, a part from the simulator sickness questionnaire (Kennedy et al., 1993), also stated their propensity to suffer motion sickness in general (yes/no) and whether they played 3D videogames (yes/no) or 2D videogames (yes/no).

Participants were recruited *via* email list. The experimental protocol employed in this study was approved by Microsoft

FIGURE 3 | **(Left) Ground station setup**. Operator flying a drone *via* our first person view system: GT control, oculus head-mounted display. (Right) The flight view from the operator perspective. Watch Movie S1 in Supplementary Material to see the setup at work.

Research, and the experimental data were collected with the approval and written consent of each participant following the Declaration of Helsinki. Voluntaries were paid with a lunch card for their participation.

Real Flight Test

To test our FPV setup for drone teleoperation, we asked seven participants (all male, mean age = 45.14, SD = 2.05) to fly our drone in the open field. Participant flew for about 15 min while wearing VR HMD. All participants had the same exposure time to reduce differences in simulator sickness (So et al., 2001). It is also known that navigation speed affects simulator sickness, being maximal at 10 m/s and stabilizing for faster navigation speeds (So et al., 2001). In our case, participants were able to achieve speeds in the range of 0–7 m/s. In the real-time condition, participants could choose the speed, but to reduce differences, they all completed the experiment in the same open space and were given same instructions as to where to go and how to land the drone on the return. Participants did not have previous experience with the setup nor with VR. After the experience, participants completed a simplified version of the simulator sickness questionnaire (Kennedy et al., 1993) that ultimately rates the four levels of sickness (none, slight, moderate, and severe).

Experiment 1

To evaluate how different viewing systems (mono vs. stereo) influence human operators' perception in a drone FPV scenario, we ran a first experiment with 20 participants (mean age = 33.75, SD = 10.18, 5 female). Participants wearing VR HMDs (Oculus Rift) were shown prerecorded stereoscopic flights and walkthroughs. We asked participants to estimate the camera height above the ground in these recordings. Half of the participants were exposed to a ground navigation and the other half to a flight navigation. All participants experienced the same video in monoscopic and stereoscopic view (in counterbalanced order) and were asked to report their perceived height in both conditions.

Experiment 2

To further study the incidence of simulator sickness, we recruited 69 participants (mean age = 33.84, SD = 9.50, 13 female) to

experience the stereoscopic flight through prerecorded videos in VR HMD. Participants were initially randomly recruited, and 61% of the initial sample had correct vision, 29% were nearsighted, and 10% were farsighted. These demographics are in line with previous studies of refractive error prevalence (The Eye Diseases Prevalence Research Group, 2004). Additional farsighted participants were recruited to achieve a larger number of participants in that category ($n = 16$). The final demographics of the study contained 52% of population with correct vision, 25% of nearsighted, and 23% farsighted. Participants in this experiment also stated their propensity to suffer motion sickness in general (yes/no), whether they played 3D videogames (yes/no), and their visual acuity (20/20, farsighted, and nearsighted). Participants were clarified that they would classify as nearsighted if they could not properly focus on distant objects; farsighted if they could see distant objects clearly, but could not see near objects in proper focus; and 20/20 if they had normal vision without corrective lenses. Half of the participants reported propensity to have motion sickness.

RESULTS

Ground Control Experiment

Participants ($n = 7$) in the live tests could take off, fly, and land the drone at speeds less than 7 m/s, and none of the participants reported moderate-to-severe simulator sickness after the experience, even though four of them reported propensity to generally suffer motion sickness and only three of them played 3D games. Users reported the sensation of flying and out-of-body experiences (Blanke and Mohr, 2005). In addition, participants showed a great tolerance for latency, since the system had 250 ± 30 ms latency, but while controlling the drone, they could not notice the latency, which was clear for an external observer. This effect might have been triggered by action-binding mechanisms, by which perceived time of intentional actions and of their sensory consequences that can affect consciousness of actions by reducing perceived latencies (Haggard et al., 2002). Overall, the system was able to achieve top speeds of 7 m/s without accidents. Even though the drone was capable of faster propulsion, the latency together with the action-binding latency perception demonstrated critical

at higher than 7 m/s speeds, speed at which landing of the drone could provoke involuntary crashes. These speeds are limited compared to the 16 m/s speed advertised by the drone manufacturer; however, that speed is maximal only with no payload, and our camera setup with batteries weighs over 3.3 kg; in fact, the manufacturer does not recommend flying at such speeds.[9]

The sickness level between the real-time and offline conditions was significantly different (Pearson's Chi-squared test of independence, $\chi^2 = 9.5$, df = 3, $p = 0.02$), and participants in the real-time condition did not report simulator sickness, while 29% of the participants in Experiment 2 reported moderate-to-severe motion sickness.

Experiment 1

An important feature for unmanned drone navigation is the ability for both the embedded autonomous system and the drone operator to perform distance estimations. In the monoscopic condition, participants significantly overestimated distances when compared to their estimation on the stereoscopic condition (Wilcoxon signed-rank paired test $V = 52$, $p = 0.013$, $n = 40$; **Figure 4**). Although the real height/altitude of the camera was 1.8 m in the Ground scene (walkthrough), participants of the monoscopic condition estimated the height/altitude to be 2.1 ± 0.16 m, and participants in the stereoscopic condition estimated 1.8 ± 0.15 m. Similarly, in the Flight scene (drone flight) where the real height/drone altitude was 3.5 m, participants in the monoscopic condition estimated the height/altitude to be 3.9 ± 0.7 m, while the estimation was more accurate in the stereoscopic condition 3.4 ± 0.6 m.

We found that participants who reported their experience to be better in the stereoscopic condition were also more accurate at the distance estimation in that condition (Pearson, $p = 0.06$, $r = 0.16$). Similarly, participants who reported their experience to be better in the monoscopic condition were also more accurate at the distance estimation during the monoscopic condition (Pearson, $p = 0.05$, $r = -0.44$). However, we did not find a significant correlation between the height/altitude perception and the IPD of the participants (Pearson, $p = 0.97$, $r = 0.0$).

In general, over 40% of the participants in this experiment reported no motion sickness at all during the experience, while 28% reported slight, 15% moderate, and 18% severe motion sickness. We did not find a significant difference across modes and conditions in simulator sickness (Wilcoxon signed-rank paired test $V = 27$, $p = 0.6078$). The difference between reported simulation sickness in the live control system and the offline experiment is mainly due to the effect of "observation only" vs. "action control."

Experiment 2

Arguably the sickness in the prerecorded experience can be considered low because 71% participants in this condition reported none-to-slight motion sickness. (**Figure 5** shows the distribution of the simulator sickness.)
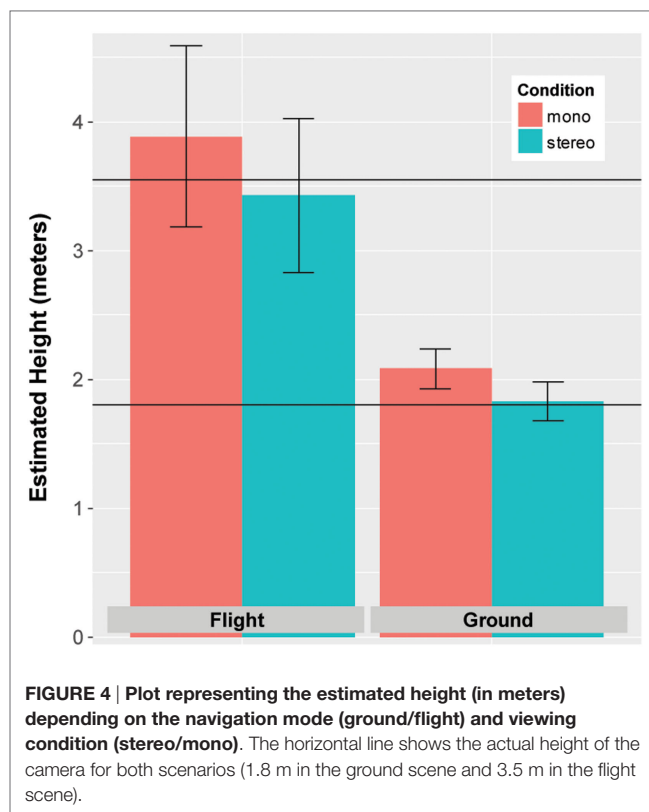
FIGURE 4 | Plot representing the estimated height (in meters) depending on the navigation mode (ground/flight) and viewing condition (stereo/mono). The horizontal line shows the actual height of the camera for both scenarios (1.8 m in the ground scene and 3.5 m in the flight scene).

To explore the covariates describing simulator sickness, we run a correlation study (**Table 1**). Results show that self-reported propensity to generally suffer of motion sickness is significantly correlated with the actual reported simulator sickness (Pearson, $p = 0.026$, $r = 0.27$); therefore, we explore the difference in simulator sickness and find it significantly higher for people with sickness propensity (Kruskal–Wallis chi-squared = 6.27, $p = 0.012$), probably indicating the importance of the predominant vestibular mechanism in the generation of simulator sickness. On the other hand, playing 3D games did not correlate with the reported motion sickness ($p = 0.98$, $r = 0.00$), i.e., visual adaptation to 3D simulated graphics cannot describe the extent of the simulator sickness when immersed in an HMD. More in particular, we find that simulator sickness was not significantly different depending on the gaming experience (yes/no) (Kruskal–Wallis chi-squared = 0.017, $p = 0.89$).

As expected, based on extensive visual acuity studies (The Eye Diseases Prevalence Research Group, 2004), we find a strong correlation of visual acuity with age ($p = 0.004$, $r = 0.34$); however, age was not a significant descriptor for motion sickness nor any other variable.

To better study the effects of simulation sickness, we divide the participants who have propensity ($n = 35$) from those who do not ($n = 34$), and we analyze the results for both groups. Participants who reported propensity to motion sickness showed different levels of sickness depending on their visual acuity (**Figure 6**). More in particular, simulator sickness was found significantly higher for farsighted participants ($n = 11$), when compared to nearsighted ($n = 8$) (Kruskal–Wallis chi-squared = 5.98,
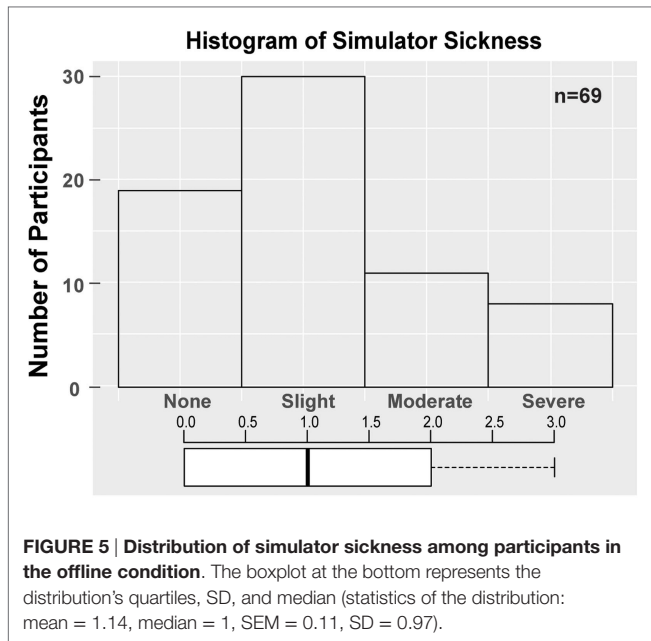
FIGURE 5 | Distribution of simulator sickness among participants in the offline condition. The boxplot at the bottom represents the distribution's quartiles, SD, and median (statistics of the distribution: mean = 1.14, median = 1, SEM = 0.11, SD = 0.97).

TABLE 1 | Pearson correlation matrix for Experiment 1.

	Simulator sickness	Age	Propensity	Vision acuity
Age	$p = 0.599$, $r = 0.06$			
Propensity	$p = \mathbf{0.026}$*, $\mathbf{r = 0.27}$	$p = 0.644$, $r = -0.06$		
Vision acuity	$p = 0.377$, $r = 0.11$	$p = \mathbf{0.004}$*, $\mathbf{r = 0.34}$	$p = 0.134$, $r = 0.18$	
3D gaming	$p = 0.983$, $r = 0.00$	$p = 0.725$, $r = -0.04$	$p = 0.736$, $r = -0.04$	$p = 0.807$, $r = 0.03$

*Significant values in bold: *$p < 0.05$.*

$p = 0.014$); no significant differences were found with normal vision ($n = 16$).

DISCUSSION

In this article, we presented a drone visual system that implements immersive stereoscopic FPV that can be used to improve current ground control scenarios. This approach can improve teleoperation and handover scenarios (Russell et al., 2016), can simplify autonomous navigation (compared to monoscopic systems), and help in capturing better video footage for training deep control policies (Artale et al., 2013).

Our results show that in the real-time condition, with our system, users can fly and land without problems. Results suggest that simulator sickness was almost non-existent for participants who could control the camera navigation in real time, but affected 29% of participants in the prerecorded condition. This is congruent with previous findings (Stanney and Hash, 1998). We hypothesize that intentional action-binding mechanisms played a key role during the real-time condition (Haggard et al., 2002;
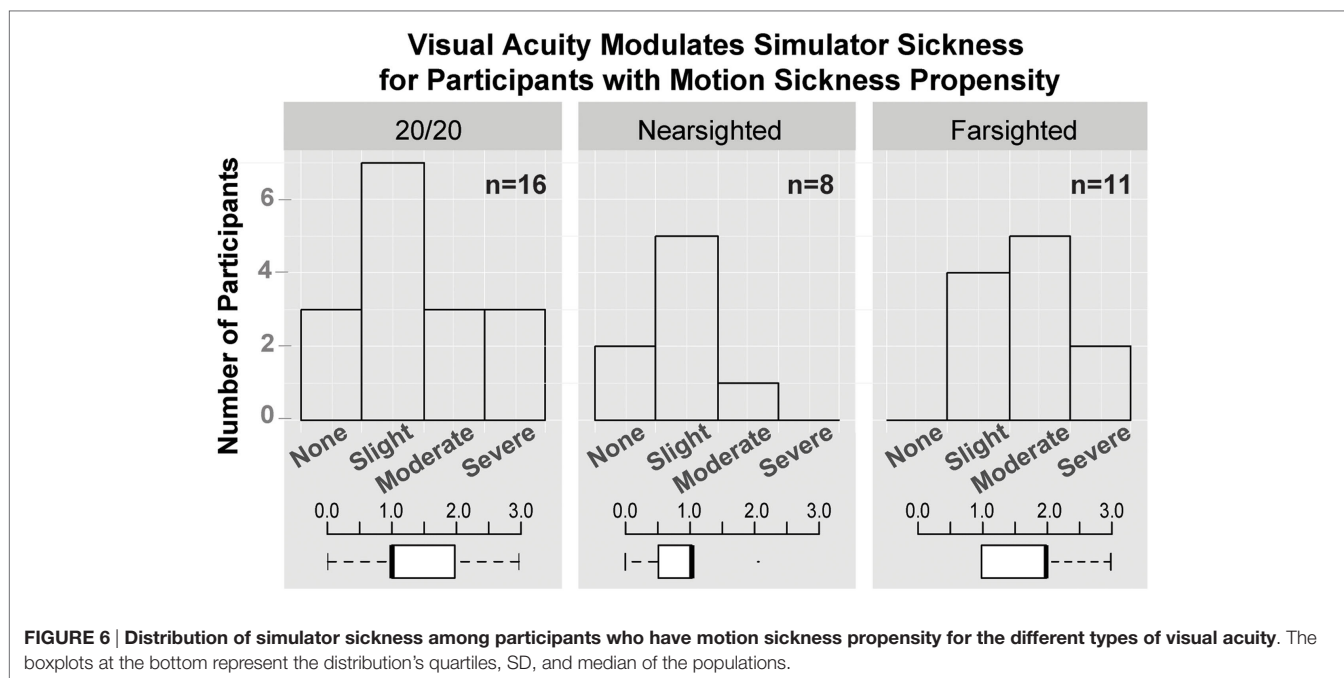
Padrao et al., 2016). Furthermore, tolerance in perceivable latencies for participants of the real-time condition is also explainable through those mechanisms (Haggard et al., 2002). In fact, having an expected outcome of the action could have helped in the offline condition, e.g., introducing visual indications of future trajectories could have reduced their simulator sickness. Further research is needed to compare our latencies to those introduced by physical panning systems; however, given the low ratios of simulator sickness during the real flights, digital panning seems to be more usable. One shortcoming of using digital panning is that depending on the view direction, it can further reduce the stereoscopic baseline. This problem would not occur with physical panning.

Our test results from Experiment 1 show that the effectiveness in flight distance estimations improved with stereoscopic vision, without significant impact on the induced motion sickness when compared to the monoscopic view, i.e., participants can guess distances and heights more accurately in the stereo setup than in mono and therefore can maneuver the drone more efficiently/safely. Hence, we believe that our system can help in providing a better experience to drone operators and improving teleoperation accuracy. A higher quality teleoperation is critical in delivering better handover experiences and in collecting better flight data for autonomous control training.

Results from Experiment 2 showed a more systematic study on the incidence of motion sickness of our immersive setup on the general population. We find that predisposition for motion sickness was a good predictor of simulator sickness, which is in agreement with previous findings (Groen and Bos, 2008). Interestingly, previous 3D gaming experience did not affect the results in our experiment. Our results also show that participants who have propensity for motion sickness ($n = 35$) had different simulator sickness distributions depending on their visual acuity. Being those with farsighted vision, the ones more affected by simulator sickness, it might be that these subjects have a higher propensity for general motion sickness due to their visual acuity. In fact, effects of visual acuity to simulator sickness are probably related to other visual imparities that have also been described to affect simulator sickness such as color blindness (Gusev et al., 2016). The fact that no significant differences were found between nearsighted and 20/20 could be related to the corrective lenses; therefore, their vision could be considered corrected to 20/20 in this experiment.

Even though we found visual acuity to correlate with age, which is in agreement with The Eye Diseases Prevalence Research Group (2004), in general (independently of propensity to motion sickness), age alone was not explanatory variable for simulator sickness.

Previous reviews on US Navy flight simulators reported that incidence of sickness varied from 10 to 60% (Barrett, 2004). The wide variation on results seems to depend on the design of the experience and the technology involved. Indeed, HMDs have evolved significantly in FOV and resolution, both aspects being critical to simulator sickness (Lin et al., 2002). In that sense, our study is relevant as it uses state-of-the-art technology and explores the prevalence for the specific experience of camera captured VR navigation during drone operation for a considerable population.

FIGURE 6 | Distribution of simulator sickness among participants who have motion sickness propensity for the different types of visual acuity. The boxplots at the bottom represent the distribution's quartiles, SD, and median of the populations.

Although other recent experiments on simulator sickness have concentrated more on computer graphic generated VR and do not feature drone operation (Fernandes and Feiner, 2016). Given the importance of simulator sickness in the spread of FPV control for drones (Biocca, 1992), and the expected growth in the consumer arena, we hope these results shed light on the perceptual tolerances during navigation of drones from an FPV using VR setups.

To summarize, our results provide evidence that an immersive stereoscopic FPV control of drones is feasible and might enhance the ground control experience. Furthermore, we provide a comprehensive study of implications at the simulator sickness level of such a setup. Our proposed FPV setup is to be understood as a piece to be combined with autonomous obstacle avoidance algorithms that run on the drone's onboard computer using the same cameras.

There are some shortcomings in our setup. On the one hand, the real flight testing was limited, and most of the study was done with prerecorded data. Nevertheless, the results on simulator sickness might be comparable. Indeed, our real flight results suggest that the system might even be performing better in terms of simulator sickness. On the other hand, we did not compare directly to TPV systems that are also very dominant in consumer drones. TPV systems do not suffer from simulator sickness and might allow for larger speeds. However, in the context of out-of-sight drones and handover situations, TPV might not always be possible, and FPV systems can solve those scenarios. In addition, we believe that further experimentation with the baselines of the cameras might be interesting. Human baseline is limited to 19 m (Allison et al., 2009), but with camera capturing systems, this can be modified and enable greater distance estimations when necessary. Our

system had also issues that would make it hard to operate in a real environment, such as the weight of the setup, 3.3 kg with batteries, which significantly reduces the remaining payload capacity of the drone and the latency of 250 ms, which makes it impossible to fly the drone at speeds higher than 7 m/s. Both aspects, weight and latency, can be improved in future development after the prototyping phase, for example, by miniaturizing the cameras and improving the streaming and encoding computation.

Overall, we believe that such a system will enhance the control and security of the drone with trajectory correction, crash prevention, and safety landing systems. Furthermore, the combination of a human FPV drone control with automatic low-level obstacle avoidance is superior to fully autonomous systems at the time of writing. In fact, current regulations are required to contemplate handover situations (Russell et al., 2016). While the human operator is responsible for higher level navigation guidance with current regulations, the low-level automatic obstacle avoidance could decrease operator's cognitive load and increase the security of the system. Our FPV system can be used for both operator's FPV drone control and automatic obstacle avoidance. In addition, our system can be used to collect training data for machine learning algorithms for fully autonomous drones. Modern machine learning approaches with deep neural networks (Artale et al., 2013) need big data sets to train the systems; such data sets need to be collected with optimal navigation examples that can be obtained using ground FPV navigation systems. With our setup, the same stereo cameras can be used to evaluate the navigation during semiautonomous or autonomous flying and also to provide an optimal fully immersive FPV drone control from the ground when needed.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

Allison, R. S., Gillam, B. J., and Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *J. Vis.* 9, 10.1–14. doi:10.1167/9.1.10

Artale, V., Collotta, M., Pau, G., and Ricciardello, A. (2013). Hexacopter trajectory control using a neural network. *AIP Conf. Proc.* 1558, 1216–1219. doi:10.1063/1.4825729

Barrett, J. (2004). *Side Effects of Virtual Environments: A Review of the Literature*. No. DSTO-TR-1419. Canberra, Australia: Defence Science and Technology Organisation.

Biocca, F. (1992). Will simulation sickness slow down the diffusion of virtual environment technology? *Presence Teleop. Virtual Environ.* 1, 334–343. doi:10.1162/pres.1992.1.3.334

Blanke, O., and Mohr, C. (2005). Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Res. Brain Res. Rev.* 50, 184–199. doi:10.1016/j.brainresrev.2005.05.008

Bowman, D. A., and McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *IEEE Computer* 40, 36–43. doi:10.1109/MC.2007.257

Chen, J. Y. C., Haas, E. C., and Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 37, 1231–1245. doi:10.1109/TSMCC.2007.905819

Drascic, D. (1991). Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 35, 1367–1371. doi:10.1177/154193129103501906

Fernandes, A. S., and Feiner, S. K. (2016). "Combating VR sickness through subtle dynamic field-of-view modification," in *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (Greenville, SC), 201–210.

Fishman, J. M., Ellis, S. R., Hasser, C. J., and Stern, J. D. (2008). Effect of reduced stereoscopic camera separation on ring placement with a surgical telerobot. *Surg. Endosc.* 22, 2396–2400. doi:10.1007/s00464-008-0032-8

Forster, C., Pizzoli, M., and Scaramuzza, D. (2014). "SVO: Fast semi-direct monocular visual odometry," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong), 15–22.

González Franco, M. (2014). *Neurophysiological Signatures of the Body Representation in the Brain Using Immersive Virtual Reality*. Universitat de Barcelona.

González-Franco, M., Pérez-Marcos, D., Spanlang, B., and Slater, M. (2010). "The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment," in *2010 IEEE Virtual Reality Conference (VR)* (Waltham, MA: IEEE), 111–114. doi:10.1109/VR.2010.5444805

Groen, E. L., and Bos, J. E. (2008). Simulator sickness depends on frequency of the simulator motion mismatch: an observation. *Presence* 17, 584–593. doi:10.1162/pres.17.6.584

Gusev, D. A., Whittinghill, D. M., and Yong, J. (2016). A simulator to study the effects of color and color blindness on motion sickness in virtual reality using head-mounted displays. *Mob. Wireless Technol.* 2016, 197–204. doi:10.1007/978-981-10-1409-3_22

Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nat. Neurosci.* 5, 382–385. doi:10.1038/nn827

Hals, E., Prescott, J., Svensson, M. K., and Wilthil, M. F. (2014). "Oculus FPV," in *TPG4850 – Expert Team, VR-Village* (Norwegian University of Science and Technology).

Hayakawa, H., Fernando, C. L., Saraiji, M. H. D., Minamizawa, K., and Tachi, S. (2015). "Telexistence drone: design of a flight telexistence system for immersive aerial sports experience," in *2015 ACM Proceedings of the 6th Augmented Human International Conference* (Singapore), 171–172.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi:10.1207/s15327108ijap0303_3

Kishore, S., González-Franco, M., Hintermuller, C., Kapeller, C., Guger, C., Slater, M., et al. (2014). Comparison of SSVEP BCI and eye tracking for controlling a humanoid robot in a social environment. *Presence Teleop. Virtual Environ.* 23, 242–252. doi:10.1162/PRES_a_00192

Lin, J. J.-W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., and Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proc. IEEE Virtual Real.* 2002, 164–171. doi:10.1109/VR.2002.996519

Lynen, S., Achtelik, M. W., Weiss, S., Chli, M., and Siegwart, R. (2013). "A robust and modular multi-sensor fusion approach applied to MAV navigation," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* (Tokyo), 3923–3929.

Mirk, D., and Hlavacs, H. (2015). "Virtual tourism with drones: experiments and lag compensation," in *2015 ACM Proceedings of the First Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use (DroNet '15)* (Florence), 45–50.

Padrao, G., Gonzalez-Franco, M., Sanchez-Vives, M. V., Slater, M., and Rodriguez-Fornells, A. (2016). Violating body movement semantics: neural signatures of self-generated and external-generated errors. *Neuroimage* 124 PA, 174–156. doi:10.1016/j.neuroimage.2015.08.022

Pausch, R., Crea, T., and Conway, M. (1992). A literature survey for virtual environments: military flight simulator visual systems and simulator sickness. *Presence* 1, 344–363. doi:10.1016/S0022-3913(12)00047-9

Pepper, R., Cole, R., and Spain, E. (1983). The influence of camera separation and head movement on perceptual performance under direct and tv-displayed conditions. *Proc. Soc. Inf. Disp.* 24, 73–80.

Pryor, H. (1969). Objective measurement of interpupillary distance. *Pediatrics* 44, 973–977.

Ramasubramanian, V. (2015). *Quadrasense: Immersive UAV-Based Cross-Reality Environmental Sensor Networks*. Massachusetts Institute of Technology.

Rosenberg, L. B. (1993). The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks. *Proc. IEEE Virtual Real. Annu. Int. Symp.* 27–32. doi:10.1109/VRAIS.1993.380802

Russell, H. E. B., Harbott, L. K., Nisky, I., Pan, S., Okamura, A. M., and Gerdes, J. C. (2016). Motor learning affects car-to-driver handover in automated vehicles. *Sci. Robot.* 1, 1. doi:10.1126/scirobotics.aah5682

Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339. doi:10.1038/nrn1651

Smets, G., and Overbeeke, K. (1995). Trade-off between resolution and interactivity in spatial task performance. *IEEE Comput. Graph. Appl.* 15, 46–51. doi:10.1109/38.403827

So, R. H., Lo, W. T., and Ho, A. T. (2001). Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Hum. Factors* 43, 452–461. doi:10.1518/001872001775898223

Spanlang, B., Normand, J.-M., Borland, D., Kilteni, K., Giannopoulos, E., Pomes, A., et al. (2014). How to build an embodiment lab: achieving body

SUPPLEMENTARY MATERIAL

MOVIE S1 | Example of drone flying with the stereoscopic FPV system proposed in the paper where an operator controls the drone immersively. The video offers views from a FPV as well as third person views of the drone taking off, flying, and landing.

representation illusions in virtual reality. *Front. Robot. AI* 1:1–22. doi:10.3389/frobt.2014.00009

Stanney, K. M., and Hash, P. (1998). Locus of user-initiated control in virtual environments: influences on cybersickness. *Presence Teleop. Virtual Environ.* 7, 447–459. doi:10.1162/105474698565848

Stoffregen, T. A., Hettinger, L. J., Haas, M. W., Roe, M. M., and Smart, L. J. (2000). Postural instability and motion sickness in a fixed-base flight simulator. *Hum. Factors J. Hum. Factors Ergon. Soc.* 42, 458–469. doi:10.1518/001872000779698097

The Eye Diseases Prevalence Research Group. (2004). The prevalence of refractive errors among adults in the United States, Western Europe, and Australia. *Arch. Ophthalmol.* 122, 495–505. doi:10.1001/archopht.122.4.495

Xiao, R., and Benko, H. (2016). "Augmenting the field-of-view of head-mounted displays with sparse peripheral displays," in *2016 ACM Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (Santa Clara, CA).

Zhang, Y., Wu, L., and Wang, S. (2013). UCAV path planning by fitness-scaling adaptive chaotic particle swarm optimization. *Math. Probl. Eng.* 2013. doi:10.1155/2013/705238

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1330–1334. doi:10.1109/34.888718

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.